

# A NEW METHOD FOR COMPUTING $\varphi$ -FUNCTIONS AND THEIR CONDITION NUMBERS OF LARGE SPARSE MATRICES

GANG WU\* AND LU ZHANG†

**Abstract.** We propose a new method for computing the  $\varphi$ -functions of large sparse matrices with low rank or fast decaying singular values. The key is to reduce the computation of  $\varphi_\ell$ -functions of a large matrix to  $\varphi_{\ell+1}$ -functions of some  $r$ -by- $r$  matrices, where  $r$  is the numerical rank of the large matrix in question. Some error analysis on the new method is given. Furthermore, we propose two novel strategies for estimating 2-norm condition numbers of the  $\varphi$ -functions. Numerical experiments illustrate the numerical behavior of the new algorithms and show the effectiveness of our theoretical results.

**Key words.** Matrix function;  $\varphi$ -functions; Fréchet derivative; Low-rank matrix; Fast decaying singular values; Sparse column-row approximation (SCR).

**AMS subject classifications.** 65F60, 65F35, 65F15.

**1. Introduction.** In recent years, a great deal of attention has been focused on the efficient and accurate evaluation of matrix functions closely related to the  $\varphi$ -functions [4, 20, 22, 23, 27, 29, 30, 33, 35, 42]. For instance, exponential integrators make use of the matrix exponential and related matrix functions within the formulations of the numerical methods, and the evaluation of matrix functions is crucial for accuracy, stability, and efficiency of exponential integrators [23]. The  $\varphi$ -functions are defined for scalar arguments by the integral representation as follows

$$\varphi_0(z) = \exp(z) \quad \text{and} \quad \varphi_\ell(z) = \frac{1}{(\ell-1)!} \int_0^1 \exp((1-\theta)z) \theta^{\ell-1} d\theta, \quad \ell = 1, 2, \dots \quad (1.1)$$

Moreover, the  $\varphi$ -functions satisfy the following recurrence relations

$$\varphi_\ell(z) = z\varphi_{\ell+1}(z) + \frac{1}{\ell!}, \quad \ell = 0, 1, 2, \dots \quad (1.2)$$

This definition can be extended to matrices instead of scalars by using any of the available definitions of matrix functions [20, 23].

In a wide range of applications, such as the matrix exponential discriminant analysis method for data dimensionality reduction [1, 11, 40, 41, 43, 44], and the complex network analysis method based on matrix function [3, 13, 14, 15], it is required to compute the matrix exponential with respect to large scale and low-rank matrix. In this paper, we are interested in computing several  $\varphi$ -functions *consecutively*, with respect to a large scale matrix  $A$  with low rank or with fast decaying singular values. Let  $\sigma_j$  be the  $j$ -th largest singular value of  $A$ , by “fast decaying singular values”, we mean  $\sigma_j = \mathcal{O}(\rho^{-j})$ ,  $\rho > 1$  or  $\sigma_j = \mathcal{O}(j^{-\alpha})$ ,  $\alpha > 1$  [24]. Such matrices appear frequently in diverse application areas such as data dimensionality reduction [12], complex network analysis [14], discretizing ill-posed operator equations that model

\*Corresponding author (G. Wu). Department of Mathematics, China University of Mining and Technology, Xuzhou, 221116, P.R. China. E-mail: [gangwu76@126.com](mailto:gangwu76@126.com) and [gangwu@cumt.edu.cn](mailto:gangwu@cumt.edu.cn). This author is supported by the National Science Foundation of China under grant 11371176, the Natural Science Foundation of Jiangsu Province under grant BK20131126, and the Talent Introduction Program of China University of Mining and Technology.

†School of Mathematics and Physical Sciences Technology, Xuzhou Institute of Technology, Xuzhou, 221111, Jiangsu, P.R. China. E-mail: [yulu7517@126.com](mailto:yulu7517@126.com).

many inverse problems [18], randomized algorithms for matrix approximations [17, 28], finite elements discretization [10], and so on.

In spite of the high demand for efficient methods to solve the matrix  $\varphi$ -functions in various fields of computational sciences, there is no easy way to solving this type of problem. Indeed, when  $A$  is large, both the computational cost and the storage requirements are prohibitive, moreover,  $\varphi_\ell(A)$  can be dense even if  $A$  is sparse [20]. Some available methods are only suitable for medium-sized matrices. For instance, a MATLAB toolbox called EXPINT is provided by Berland, Skaflestad and Wright [4] for this problem. Kassam and Trefethen [26] propose to approximate  $\varphi$ -functions with a contour integral, which worked well as long as the contour of integration is suitably chosen. However, the contour is in general problem-dependent and difficult to determine in advance. Another way is to reduce the computation of  $\varphi_\ell(A)$  to that of matrix exponential with larger size [2, 27, 34], which is unfeasible as the matrix in question is large. The third way is based on a modification of the scaling and squaring technique [35], the most commonly used approach for computing the matrix exponential [21]. The most well-known methods for computing  $\varphi$ -functions for large sparse matrices are the Krylov subspace methods [30, 42]. However, the Krylov subspace methods are applicable to the computation of  $\varphi$ -functions on given (block) vectors, while the main aim of this paper is to compute  $\varphi$ -functions of large sparse matrices.

In practical calculations, it is important to know how accurate the computed solution is and how small perturbations in input data can effect outputs [19]. Therefore, it is crucial to give error analysis and understand the sensitivity of matrix function to perturbation in the data. Sensitivity is measured by condition numbers. For matrix function, condition number can be expressed in terms of the norm of the Fréchet derivative, and it is often measured by using the 1-norm [20, 46]. In practice, however, the 2-norm is a more widely used norm than the 1-norm, and the former is preferable for both theoretical analysis and computational purposes. In this work, we consider how to evaluate the Fréchet 2-norm condition numbers of  $\varphi$ -functions effectively.

Given a large scale matrix  $A$  with low rank or with fast decaying singular values, we propose a new method for evaluating several  $\varphi$ -functions and their absolute and relative condition numbers *consecutively*. Our new method is based on the sparse column-row approximation of large sparse matrices [8, 36, 37]. An advantage is that there is no need to explicitly form and store the  $\varphi$ -functions or the Fréchet derivatives with respect to  $A$ . The overhead is only to compute  $\varphi$ -functions of some  $r$ -by- $r$  matrices, and to store two  $n$ -by- $r$  sparse matrices, where  $r$  is the (numerical) rank of  $A$ . This paper is organized as follows. In Section 2, we present the main algorithm, and give some error analysis on the proposed method. In Section 3, we propose two novel strategies for estimating the absolute and relative 2-norm condition numbers of  $\varphi$ -functions. In Section 4, numerical experiments are given to illustrate the efficiency of our new strategies. Some concluding remarks are given in Section 5.

Some notations used are listed as follows. Throughout this paper, we denote by  $\tilde{A} = XTY^T$  a sparse column-row approximation to  $A$ . Let  $\|\cdot\|_2, \|\cdot\|_1$  be the 2-norm and the 1-norm of a vector or matrix, and  $\|\cdot\|_F$  be the Frobenius norm of a matrix. We denote by  $\otimes$  the Kronecker product, and by  $\text{vec}(\cdot)$  the “vec operator” that stacks the columns of a matrix into a long vector. Let  $I$  and  $O$  be the identity matrix and zero matrix, respectively, whose order is clear from context. We only focus on real matrices in this paper. Indeed, all the results can be extended to complex matrices in a similar way.

**2. A new method for  $\varphi$ -functions of large sparse matrices.** In this section, we will present a new method for  $\varphi$ -functions of large sparse matrices with low rank or with fast decaying singular values, and give some error analysis on the proposed method. Given an  $n \times n$  large sparse matrix  $A$ , we first find a reduced-rank approximation  $XTY^T$  to  $A$ , where both  $X$  and  $Y$  are full column rank and  $T$  is nonsingular. This type of problem arises in a number of applications such as information retrieval, computational biology and complex network analysis [5, 6, 7, 8, 25, 38, 45]. A widely used reduced-rank approximation is the truncated singular value decomposition (TSVD) [16], which is known to be optimal in the sense that the Frobenius norm  $\|A - XTY^T\|_F$  is minimized. Unfortunately, this method computes the full decomposition and is not suitable for very large matrices. An alternative is the randomized singular value decomposition algorithm [17, 28], which generally gives results comparable to TSVD. However, for a large and sparse matrix  $A$ , the situation is not so simple: the storage requirements and operation counts will become proportional to the number of nonzero elements in  $A$ . Since the resulting factors  $X, T$ , and  $Y$  are generally not sparse, one may suffer from heavily computational cost.

In [36], Stewart introduced a quasi-Gram-Schmidt algorithm that produces a sparse QR factorization to  $A$ . Based on the quasi-Gram-Schmidt algorithm, a sparse column-row approximation algorithm was proposed. This algorithm first applies the quasi-Gram-Schmidt algorithm to the columns of  $A$  to get a representative set of columns  $X$  of  $A$  and an upper triangular matrix  $R$ . Let the error in the corresponding reduced-rank decomposition be  $\epsilon_{\text{col}}$ . It then applies the same algorithm to  $A^T$  to get a representative set  $Y^T$  of rows and another upper triangular matrix  $S$ . Let the error be  $\epsilon_{\text{row}}$ . Then the sparse column-row approximation method seeks a matrix  $T$  such that  $\|A - XTY^T\|_F^2 = \min$ , and the minimizer turns out to be [8, 36]

$$T = R^{-1}R^{-T}(X^TAY)S^{-1}S^{-T},$$

moreover, we have [36]

$$\|A - XTY^T\|_F^2 \leq \epsilon_{\text{col}}^2 + \epsilon_{\text{row}}^2. \quad (2.1)$$

The matrix  $XTY^T$  is called a sparse column-row approximation (SCR) to  $A$ , where  $X, Y \in \mathbb{R}^{n \times r}$  are sparse and full column rank,  $T \in \mathbb{R}^{r \times r}$  is nonsingular, and  $r$  is the (numerical) rank of  $A$ . In this approximation,  $X$  consists of a selection of the columns of  $A$ , and  $Y$  consists of a selection of the rows of  $A$ , so that when  $A$  is sparse so are both  $X$  and  $Y$ . An error analysis of the quasi-Gram-Schmidt algorithm is given in [37]. One is recommended to see [8, 36] for more details on this algorithm.

Given *any* rank-revealing decomposition of  $A$ , the following theorem shows that the computation of  $\varphi_\ell(A)$  can be reduced to that of  $\varphi_{\ell+1}$  function of an  $r \times r$  matrix, where  $r$  is the (numerical) rank of  $A$ .

**THEOREM 2.1.** *Let  $XTY^T \in \mathbb{R}^{n \times n}$  be a rank-revealing decomposition of  $A$ , where  $X, Y \in \mathbb{R}^{n \times r}$  and  $T \in \mathbb{R}^{r \times r}$ . Denote  $Z = T(Y^TX) \in \mathbb{R}^{r \times r}$ , then*

$$\varphi_\ell(XTY^T) = \frac{1}{\ell!}I + X[\varphi_{\ell+1}(Z)T]Y^T, \quad \ell = 0, 1, \dots \quad (2.2)$$

*Proof.* It follows from the definition of  $\varphi$ -functions that

$$\begin{aligned}
\varphi_\ell(XTY^T) &= \sum_{k=\ell}^{\infty} \frac{(XTY^T)^{k-\ell}}{k!} \\
&= \frac{1}{\ell!} I + \sum_{k=\ell+1}^{\infty} \frac{X(TY^T X)^{k-\ell-1} TY^T}{k!} \\
&= \frac{1}{\ell!} I + X \left( \sum_{k=\ell+1}^{\infty} \frac{(TY^T X)^{k-(\ell+1)}}{k!} \right) TY^T \\
&= \frac{1}{\ell!} I + X [\varphi_{\ell+1}(TY^T X) T] Y^T.
\end{aligned}$$

□

Let  $\tilde{A} = XTY^T$  be a sparse column-row approximation to  $A$ , then we make use of  $\varphi_\ell(\tilde{A})$  as an approximation to  $\varphi_\ell(A)$ . The following algorithm can be used to compute several  $\varphi$ -functions of large sparse matrices with low rank or fast decaying singular values *consecutively*.

**ALGORITHM 1. An algorithm for computing  $\varphi$ -functions of large sparse matrices with low rank or fast decaying singular values**

1. Compute a reduced-rank approximation  $XTY^T$  to  $A$  by using, say, the sparse column-row approximation (SCR) algorithm;
2. Compute  $\varphi$ -functions of small-sized matrices:  $\varphi_{\ell+1}(TY^T X)$ ,  $\ell = 0, 1, \dots, p$ ;
3. Store  $X, Y, T$  and  $\varphi_{\ell+1}(TY^T X)$  for  $\varphi_\ell(\tilde{A})$ ,  $\ell = 0, 1, \dots, p$ . If desired, form  $\varphi_\ell(\tilde{A})$  in terms of (2.2) and use them as approximations to  $\varphi_\ell(A)$   $\ell = 0, 1, \dots, p$ .

**REMARK 2.1.** Obviously, an advantage of the proposed method is its simplicity. In conventional methods, one has to pay  $\mathcal{O}((p+1)n^3)$  flops for the computation of  $\varphi_\ell(A)$  [4, 9, 20, 35]. Given a sparse reduced-rank approximation to  $A$ , Theorem 2.1 reduces the computation of  $\varphi_\ell(A)$  to that of  $\varphi_{\ell+1}$  functions with respect to the  $r$ -by- $r$  matrix  $Z = T(Y^T X)$ , in  $\mathcal{O}((p+1)r^3)$  flops. For storage, it only needs to store two  $n$ -by- $r$  sparse matrices  $X, Y$ , and some small matrices of size  $r$ -by- $r$ , rather than the  $n$ -by- $n$  possibly dense matrices  $\varphi_\ell(A)$ ,  $\ell = 0, 1, \dots, p$ . Thus, the new method can compute  $\varphi_\ell(A)$ ,  $\ell = 0, 1, \dots, p$ , consecutively and reduce the computational complexities significantly as  $r \ll n$ .

In practice, most data are inexact or uncertain. Indeed, even if the data were exact, the computations will subject to rounding errors. So it is important to give error analysis and understand the sensitivity of matrix function to perturbation in the data. Sensitivity is measured by condition numbers. For matrix function, condition number can be expressed in terms of the norm of the Fréchet derivative. The Fréchet derivative  $L_f(A, E)$  of a matrix function  $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$  at a point  $A \in \mathbb{R}^{n \times n}$  is a linear mapping such that for all  $E \in \mathbb{R}^{n \times n}$  [20, pp.56]

$$f(A + E) - f(A) - L_f(A, E) = o(\|E\|). \quad (2.3)$$

The absolute and relative condition numbers of a matrix function  $f(A)$  are defined as [20, 32]

$$\text{cond}_{\text{abs}}(f, A) = \|L_f(A)\| = \max_{E \neq O} \frac{\|L_f(A, E)\|}{\|E\|}, \quad (2.4)$$

and

$$\text{cond}_{\text{rel}}(f, A) = \frac{\|L_f(A)\| \|A\|}{\|f(A)\|}, \quad (2.5)$$

respectively. Theoretically, the Fréchet derivative can be obtained from applying any existing methods for computing the matrix function of a  $2n \times 2n$  matrix [20]

$$f\left(\begin{bmatrix} A & E \\ O & A \end{bmatrix}\right) = \begin{bmatrix} f(A) & L_f(A, E) \\ O & f(A) \end{bmatrix}, \quad (2.6)$$

see, [20, Algorithm 3.17]. However, it requires  $\mathcal{O}(n^5)$  flops, assuming that the computation of  $L_f(A, E)$  takes  $\mathcal{O}(n^3)$  flops [20], which are prohibitively expensive for large matrices.

Let  $\tilde{A} = XTY^T$  be a sparse column-row approximation to  $A$ , and let  $E = A - \tilde{A}$ , then we see from (2.1) that  $\|E\|_F \leq \sqrt{\epsilon_{\text{col}}^2 + \epsilon_{\text{row}}^2}$ . Thus, it is interesting to combine existing error analysis for sparse column-row approximation with the theory of matrix functions to obtain error bounds for the proposed method. We first present the following theorem for Fréchet derivatives of  $\varphi$ -functions.

**THEOREM 2.2.** *For any matrix  $E \in \mathbb{R}^{n \times n}$ , we have*

$$\varphi_\ell(A + E) - \varphi_\ell(A) = \int_0^1 \exp((1-s)A) s^\ell E \varphi_\ell(s(A + E)) ds, \quad \ell = 0, 1, 2, \dots \quad (2.7)$$

*Proof.* For the matrix exponential, it follows from [20, pp.238] that

$$\exp(A + E) = \exp(A) + \int_0^1 \exp((1-s)A) E \exp(s(A + E)) ds, \quad (2.8)$$

so (2.7) holds for  $\ell = 0$ . When  $\ell \geq 1$ , let  $B = \begin{bmatrix} A & I \\ O & O \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$ , then we get [20, 33]

$$\varphi_\ell(B) = \begin{bmatrix} \varphi_\ell(A) & \varphi_{\ell+1}(A) \\ O & \frac{1}{\ell!} I \end{bmatrix}.$$

Denote  $\tilde{E} = \begin{bmatrix} E & O \\ O & O \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$ , it is seen that

$$\varphi_\ell(B + \tilde{E}) - \varphi_\ell(B) = \begin{bmatrix} \varphi_\ell(A + E) - \varphi_\ell(A) & \varphi_{\ell+1}(A + E) - \varphi_{\ell+1}(A) \\ O & O \end{bmatrix}. \quad (2.9)$$

By induction, we assume that

$$\varphi_\ell(B + \tilde{E}) - \varphi_\ell(B) = \int_0^1 \exp((1-s)B) s^\ell \tilde{E} \varphi_\ell(s(B + \tilde{E})) ds. \quad (2.10)$$

From

$$(1-s)B \exp((1-s)B) = \exp((1-s)B)(1-s)B,$$

and

$$s(B + \tilde{E}) \varphi_\ell(s(B + \tilde{E})) = \varphi_\ell(s(B + \tilde{E})) s(B + \tilde{E}),$$

we obtain

$$\exp((1-s)B) = \begin{bmatrix} \exp((1-s)A) & (1-s)\varphi_1((1-s)A) \\ O & I \end{bmatrix},$$

and

$$\varphi_\ell(s(B + \tilde{E})) = \begin{bmatrix} \varphi_\ell(s(A + E)) & s\varphi_{\ell+1}(s(A + E)) \\ O & \frac{1}{\ell!}I \end{bmatrix}.$$

Thus,

$$\begin{aligned} & \exp((1-s)B)s^\ell \tilde{E} \varphi_\ell(s(B + \tilde{E})) \\ &= \begin{bmatrix} \exp((1-s)A)s^\ell E \varphi_\ell(s(A + E)) & \exp((1-s)A)s^{\ell+1} E \varphi_{\ell+1}(s(A + E)) \\ O & O \end{bmatrix}. \end{aligned}$$

Furthermore, (2.10) can be rewritten as

$$\begin{aligned} & \varphi_\ell(B + \tilde{E}) - \varphi_\ell(B) \\ &= \begin{bmatrix} \int_0^1 \exp((1-s)A)s^\ell E \varphi_\ell(s(A + E))ds & \int_0^1 \exp((1-s)A)s^{\ell+1} E \varphi_{\ell+1}(s(A + E))ds \\ O & O \end{bmatrix}. \end{aligned} \quad (2.11)$$

From (2.9) and (2.11), we have

$$\varphi_{\ell+1}(A + E) - \varphi_{\ell+1}(A) = \int_0^1 \exp((1-s)A)s^{\ell+1} E \varphi_{\ell+1}(s(A + E))ds,$$

which completes the proof.  $\square$

Using (2.7) to substitute for  $\varphi_\ell(s(A + E))$  inside the integral, we get

$$\varphi_\ell(A + E) = \varphi_\ell(A) + \int_0^1 \exp((1-s)A)s^\ell E \varphi_\ell(sA)ds + \mathcal{O}(\|E\|^2), \quad \ell = 0, 1, 2, \dots \quad (2.12)$$

Combining with (2.3), we present the following definition for the Fréchet derivatives of  $\varphi$ -functions:

**DEFINITION 2.3.** *The Fréchet derivatives of  $\varphi$ -functions at  $A$  in the direction  $E$  is given by*

$$L_{\varphi_\ell}(A, E) = \int_0^1 \exp((1-s)A)s^\ell E \varphi_\ell(sA)ds, \quad \ell = 0, 1, 2, \dots \quad (2.13)$$

As a result,

$$\varphi_\ell(A + E) = \varphi_\ell(A) + L_{\varphi_\ell}(A, E) + o(\|E\|). \quad (2.14)$$

In summary, the following theorem shows that the values of  $\epsilon_{\text{col}}$  and  $\epsilon_{\text{row}}$  used during the sparse column-row approximation will have a direct impact upon the final accuracy of computing  $\varphi_\ell(A)$ . Note that  $XTY^T$  can be *any* low-rank approximation to  $A$  in this theorem.

**THEOREM 2.4.** *Let  $XTY^T$  be a sparse column-row approximation to  $A$ , and  $\|A - XTY^T\| = \varepsilon$ . Then we have*

$$\|\varphi_\ell(A) - \varphi_\ell(XTY^T)\| \lesssim \text{cond}_{\text{abs}}(\varphi_\ell, A) \varepsilon, \quad (2.15)$$

and

$$\frac{\|\varphi_\ell(A) - \varphi_\ell(XTY^T)\|}{\|\varphi_\ell(A)\|} \lesssim \text{cond}_{\text{rel}}(\varphi_\ell, A) \frac{\varepsilon}{\|A\|}, \quad (2.16)$$

where  $\lesssim$  represents omitting the high order term  $o(\|E\|)$ .

*Proof.* Let  $E = A - XTY^T$ , then we get from (2.14) that

$$\begin{aligned}\|\varphi_\ell(A) - \varphi_\ell(XTY^T)\| &\lesssim \|L_{\varphi_\ell}(A)\| \|E\| = \|L_{\varphi_\ell}(A)\| \varepsilon \\ &= \text{cond}_{\text{abs}}(\varphi_\ell, A) \varepsilon,\end{aligned}$$

in which the high order term  $o(\|E\|)$  is omitted. The upper bound (2.16) of the relative error is derived from (2.5) and (2.15).  $\square$

**3. New strategies for estimating the absolute and relative 2-norm condition numbers.** By Theorem 2.4, it is crucial to consider how to estimate the absolute and relative condition numbers  $\text{cond}_{\text{abs}}(\varphi_\ell, A)$  and  $\text{cond}_{\text{rel}}(\varphi_\ell, A)$  efficiently. Notice that

$$\text{vec}(ACB) = (B^T \otimes A) \text{vec}(C). \quad (3.1)$$

Since  $L_f$  is a linear operator, we have

$$\text{vec}(L_f(A, E)) = K_f(A) \text{vec}(E) \quad (3.2)$$

for some  $K_f(A) \in \mathbb{R}^{n^2 \times n^2}$  that is independent of  $E$ . The matrix  $K_f(A)$  is referred to as the Kronecker form of the Fréchet derivative [20, pp.60]. Specifically, we have

$$\|L_f(A)\|_F = \lambda_{\max}(K_f(A)^T K_f(A))^{1/2} = \|K_f(A)\|_2. \quad (3.3)$$

To estimate  $\|K_f(A)\|_2$ , the power method can be applied [20, Algorithm 3.20]. However, the power method lacks convergence tests, and because of its linear convergence rate the number of iteration required is unpredictable. Alternatively, the condition number is based on the 1-norm [20, Algorithm 3.22]. Although there is no analogue to the relation (3.3) for the 1-norm, the next result gives a relation between  $\|K_f(A)\|_1$  and  $\|L_f(A)\|_1$ .

**THEOREM 3.1.** [20, Lemma 3.18] *For  $A \in \mathbb{R}^{n \times n}$  and any function  $f$ ,*

$$\frac{\|L_f(A)\|_1}{n} \leq \|K_f(A)\|_1 \leq n \|L_f(A)\|_1. \quad (3.4)$$

In practice, however, the 2-norm is a more widely used norm than the 1-norm, and the former is preferable for both theoretical analysis and computational purposes. For example, one of the most important properties of 2-norm is the unitary invariance [16]. Thus, we focus on the 2-norm condition number instead of the 1-norm condition number in this section. The following theorem establishes a relationship between  $\|K_f(A)\|_2$  and  $\|L_f(A)\|_2$ .

**THEOREM 3.2.** *For  $A \in \mathbb{R}^{n \times n}$  and any function  $f$ ,*

$$\frac{\|L_f(A)\|_2}{\sqrt{n}} \leq \|K_f(A)\|_2 \leq \sqrt{n} \|L_f(A)\|_2. \quad (3.5)$$

*Proof.* For any  $M \in \mathbb{R}^{n \times n}$ , we have

$$\|M\|_2 \leq \|M\|_F = \|\text{vec}(M)\|_2 \leq \sqrt{n} \|M\|_2. \quad (3.6)$$

Hence, it is seen from (3.2) and (3.6) that

$$\frac{\|K_f(A) \text{vec}(E)\|_2}{\|\text{vec}(E)\|_2} = \frac{\|\text{vec}(L_f(A, E))\|_2}{\|\text{vec}(E)\|_2} \leq \frac{\sqrt{n} \|L_f(A, E)\|_2}{\|E\|_2}, \quad \forall E \in \mathbb{R}^{n \times n}, E \neq O. \quad (3.7)$$

Similarly,

$$\frac{\|K_f(A)\text{vec}(E)\|_2}{\|\text{vec}(E)\|_2} \geq \frac{\|L_f(A, E)\|_2}{\sqrt{n}\|E\|_2}, \quad \forall E \in \mathbb{R}^{n \times n}, E \neq O. \quad (3.8)$$

Maximizing over all  $E$  for (3.7) and (3.8) yields (3.5).  $\square$

Compared (3.5) with (3.4), we see that investigating the 2-norm condition number of  $K_f(A)$  is preferable to investigating its 1-norm condition number. We are ready to show how to efficiently evaluate the Fréchet 2-condition numbers of  $\varphi$ -functions for large sparse, low-rank matrices or matrices with fast decaying singular values. Two novel strategies are proposed to evaluate the absolute and relative condition numbers.

**Strategy I.** The key idea of the first strategy is to relate  $L_{\varphi_\ell}(A)$  to  $\varphi_1(Z)$  and  $\varphi_{\ell+1}(Z)$ . We notice from (2.13) and (3.1) that

$$\begin{aligned} \text{vec}(L_{\varphi_\ell}(A, E)) &= \int_0^1 \text{vec}(\exp((1-s)A)s^\ell E \varphi_\ell(sA)) ds \\ &= \int_0^1 (\varphi_\ell(sA^T) \otimes \exp((1-s)A)s^\ell) \text{vec}(E) ds \\ &= (I \otimes \exp(A)) \int_0^1 (\varphi_\ell(sA^T) \otimes \exp(-sA)) s^\ell ds \text{vec}(E). \end{aligned} \quad (3.9)$$

Let  $X = Q_1 R_1, Y = Q_2 R_2$  be the (sparse) QR decomposition of  $X$  and  $Y$ , respectively, where  $Q_1, Q_2 \in \mathbb{R}^{n \times r}$  are orthonormal and  $R_1, R_2 \in \mathbb{R}^{r \times r}$  are upper triangular. Motivated by Theorem 2.1 and (3.9), in Strategy I we make use of

$$\text{cond}_{\text{abs}}^I(\varphi_\ell, \tilde{A}) = \|R_1 \varphi_1(Z) T R_2^T \cdot R_1 \varphi_{\ell+1}(Z) T R_2^T\|_2 \quad (3.10)$$

as an estimation to the absolute condition number  $\text{cond}_{\text{abs}}(\varphi_\ell, A)$ .

Theorem 2.1 also provides a cheap way to estimate 2-norms of  $\varphi_\ell(A)$ ,  $\ell = 0, 1, \dots, p$ . Indeed, we have from (2.2) that

$$\left| \|R_1[\varphi_{\ell+1}(Z)T]R_2^T\|_2 - 1/\ell! \right| \leq \|\varphi_\ell(\tilde{A})\|_2 \leq 1/\ell! + \|R_1[\varphi_{\ell+1}(Z)T]R_2^T\|_2. \quad (3.11)$$

Thus, we can use

$$\eta_\ell = \|R_1[\varphi_{\ell+1}(Z)T]R_2^T\|_2, \quad \ell = 0, 1, \dots, p, \quad (3.12)$$

as approximations to  $\|\varphi_\ell(A)\|_2$ . In view of (3.12), the relative condition number  $\text{cond}_{\text{rel}}(\varphi_\ell, A)$  can be approximated by using

$$\text{cond}_{\text{rel}}^I(\varphi_\ell, \tilde{A}) = \frac{\|A\|_2 \|R_1 \varphi_1(Z) T R_2^T \cdot R_1 \varphi_{\ell+1}(Z) T R_2^T\|_2}{\|R_1[\varphi_{\ell+1}(Z)T]R_2^T\|_2}. \quad (3.13)$$

Recall that there is no need to form and store the Q-factors  $Q_1$  and  $Q_2$  in practice.

**Strategy II.** The key idea of the second strategy is to relate  $L_{\varphi_\ell}(A)$  to  $L_{\varphi_{\ell+1}}(Z)$ . Recall that  $\varphi_\ell(z)$  can be expressed as the following power series whose radius of convergence is  $\infty$ :

$$\varphi_\ell(z) = \sum_{i=0}^{\infty} \frac{z^i}{(i+\ell)!}.$$



The following proposition can be viewed as a generalization of Theorem 2.1 to any matrix function in power series.

PROPOSITION 3.1. *Suppose that the power series  $f(z) = \sum_{i=0}^{\infty} \alpha_i z^i$  has radius of convergence  $\rho$ . Let  $\tilde{A} = XTY^T \in \mathbb{R}^{n \times n}$ , where  $X, Y \in \mathbb{R}^{n \times r}$  and  $T \in \mathbb{R}^{r \times r}$ . Let  $g(z) = \sum_{i=1}^{\infty} \alpha_i z^{i-1}$  and suppose that  $\|\tilde{A}\| < \rho$ , then*

$$f(\tilde{A}) = f(O) + Xg(Z)TY^T,$$

where  $Z = T(Y^T X) \in \mathbb{R}^{r \times r}$ .

*Proof.* It is seen that  $\tilde{A}^k = XZ^{k-1}TY^T$ ,  $k \geq 1$ . Thus,

$$\begin{aligned} f(\tilde{A}) &= \alpha_0 I + \alpha_1 \tilde{A} + \cdots + \alpha_k \tilde{A}^k + \cdots \\ &= \alpha_0 I + X[\alpha_1 I + \alpha_2 Z + \cdots + \alpha_k Z^{k-1} + \cdots]TY^T \\ &= f(O) + Xg(Z)TY^T. \end{aligned}$$

□

The following theorem gives closed-form formulae for  $K_f(\tilde{A})$  and  $K_g(Z)$ .

THEOREM 3.3. *Under the above notations, we have*

$$K_g(Z) = \sum_{i=2}^{\infty} \alpha_i \sum_{j=1}^{i-1} ((Z^T)^{i-j-1} \otimes Z^{j-1}). \quad (3.14)$$

Denote  $W = YT^T$ , then

$$K_f(\tilde{A}) = \Psi_1 + \Psi_2 + \Psi_3, \quad (3.15)$$

where

$$\Psi_1 = \alpha_1 I \otimes I,$$

$$\Psi_2 = (W \otimes I) \sum_{i=2}^{\infty} \alpha_i ((Z^T)^{i-2} \otimes I) (X \otimes I)^T + (I \otimes X) \sum_{i=2}^{\infty} \alpha_i (I \otimes Z^{i-2}) (I \otimes W)^T,$$

and

$$\Psi_3 = (W \otimes X) \left( \sum_{i=3}^{\infty} \alpha_i \sum_{j=2}^{i-1} ((Z^T)^{i-j-1} \otimes Z^{j-2}) \right) (X \otimes W)^T.$$

*Proof.* It follows from (2.6) and the expression of  $g(x)$  that

$$L_g(Z, F) = \sum_{i=2}^{\infty} \alpha_i \sum_{j=1}^{i-1} Z^{j-1} F Z^{i-j-1}, \quad \forall F \in \mathbb{R}^{r \times r}.$$

By (3.1),

$$\begin{aligned} \text{vec}(L_g(Z, F)) &= \sum_{i=2}^{\infty} \alpha_i \sum_{j=1}^{i-1} \text{vec}(Z^{j-1} F Z^{i-j-1}) \\ &= \sum_{i=2}^{\infty} \alpha_i \sum_{j=1}^{i-1} ((Z^T)^{i-j-1} \otimes Z^{j-1}) \text{vec}(F), \end{aligned}$$

so we get (3.14). Similarly, for any  $E \in \mathbb{R}^{n \times n}$ , we have

$$\begin{aligned}
L_f(\tilde{A}, E) &= \sum_{i=1}^{\infty} \alpha_i \sum_{j=1}^i \tilde{A}^{j-1} E \tilde{A}^{i-j} \\
&= \alpha_1 E + \alpha_2 (E \tilde{A} + \tilde{A} E) + \sum_{i=3}^{\infty} \alpha_i \left( E \tilde{A}^{i-1} + \sum_{j=2}^{i-1} \tilde{A}^{j-1} E \tilde{A}^{i-j} + \tilde{A}^{i-1} E \right) \\
&= \alpha_1 E + \sum_{i=2}^{\infty} \alpha_i (E \tilde{A}^{i-1} + \tilde{A}^{i-1} E) + \sum_{i=3}^{\infty} \alpha_i \sum_{j=2}^{i-1} \tilde{A}^{j-1} E \tilde{A}^{i-j}. \tag{3.16}
\end{aligned}$$

As a result,

$$\text{vec}(\alpha_1 E) = (\alpha_1 I \otimes I) \text{vec}(E) = \Psi_1 \text{vec}(E), \tag{3.17}$$

and we have from  $\tilde{A}^{i-1} = X Z^{i-2} T Y^T = X Z^{i-2} W^T$  ( $i \geq 2$ ) that

$$\begin{aligned}
\text{vec}\left(\sum_{i=2}^{\infty} \alpha_i (E \tilde{A}^{i-1} + \tilde{A}^{i-1} E)\right) &= \text{vec}\left(\sum_{i=2}^{\infty} \alpha_i (E X Z^{i-2} W^T + X Z^{i-2} W^T E)\right) \\
&= \sum_{i=2}^{\infty} \alpha_i (W (Z^T)^{i-2} X^T \otimes I + I \otimes X Z^{i-2} W^T) \text{vec}(E) \\
&= \left( \sum_{i=2}^{\infty} \alpha_i (W \otimes I) ((Z^T)^{i-2} \otimes I) (X \otimes I)^T \right. \\
&\quad \left. + \sum_{i=2}^{\infty} \alpha_i (I \otimes X) (I \otimes Z^{i-2}) (I \otimes W)^T \right) \text{vec}(E) \\
&= \Psi_2 \text{vec}(E). \tag{3.18}
\end{aligned}$$

Moreover,

$$\begin{aligned}
\text{vec}\left(\sum_{i=3}^{\infty} \alpha_i \sum_{j=2}^{i-1} \tilde{A}^{j-1} E \tilde{A}^{i-j}\right) &= \sum_{i=3}^{\infty} \alpha_i \sum_{j=2}^{i-1} \text{vec}(X Z^{j-2} W^T \cdot E \cdot X Z^{i-j-1} W^T) \\
&= \sum_{i=3}^{\infty} \alpha_i \sum_{j=2}^{i-1} ((W (Z^T)^{i-j-1} X^T) \otimes (X Z^{j-2} W^T)) \text{vec}(E) \\
&= \sum_{i=3}^{\infty} \alpha_i \sum_{j=2}^{i-1} (W \otimes X) ((Z^T)^{i-j-1} \otimes Z^{j-2}) (X \otimes W)^T \text{vec}(E) \\
&= \Psi_3 \text{vec}(E), \tag{3.19}
\end{aligned}$$

and (3.15) follows from (3.16)–(3.19).  $\square$

**REMARK 3.1.** *Theorem 3.3 indicates that  $L_f(\tilde{A})$  and  $L_g(Z)$  are closely related. More precisely, let*

$$\phi_i(Z) = \sum_{j=2}^{i-1} ((Z^T)^{i-j-1} \otimes Z^{j-2}), \quad i = 3, 4, \dots$$

then  $\Psi_3 = (W \otimes X) \sum_{i=3}^{\infty} \alpha_i \phi_i(Z) (X \otimes W)^T$ . On the other hand, if we denote

$$\psi_i(Z) = \sum_{j=1}^{i-1} ((Z^T)^{i-j-1} \otimes Z^{j-1}), \quad i = 2, 3, \dots$$

then  $K_g(Z) = \sum_{i=2}^{\infty} \alpha_i \psi_i(Z)$ , and it is seen from (3.14) that

$$\psi_{i-1}(Z) = \phi_i(Z), \quad i = 3, 4, \dots$$

Let  $\varphi_\ell(\tilde{A}) = \frac{1}{\ell!} I + X [\varphi_{\ell+1}(Z)T] Y^T$  and let  $\widetilde{\varphi_\ell(\tilde{A})} = \frac{1}{\ell!} I + X [\varphi_{\ell+1}(Z+F)T] Y^T$ . Then we have from (2.3) that

$$L_{\varphi_{\ell+1}}(Z, F) = \varphi_{\ell+1}(Z+F) - \varphi_{\ell+1}(Z) + o(\|F\|), \quad \forall F \in \mathbb{R}^{r \times r}.$$

Hence,

$$\begin{aligned} \widetilde{\varphi_\ell(\tilde{A})} - \varphi_\ell(A) &= X [\varphi_{\ell+1}(Z+F) - \varphi_{\ell+1}(Z)] T Y^T \\ &= X L_{\varphi_{\ell+1}}(Z, F) T Y^T + o(\|F\|). \end{aligned}$$

Let  $X = Q_1 R_1, Y = Q_2 R_2$  be the (sparse) QR decomposition of  $X$  and  $Y$ , respectively, where  $R_1, R_2 \in \mathbb{R}^{r \times r}$ . Inspired by Theorem 3.3, in Strategy II we make use of

$$\text{cond}_{\text{abs}}^{\text{II}}(\varphi_\ell, \tilde{A}) = \|X L_{\varphi_{\ell+1}}(Z) T Y^T\|_2 = \|R_1 L_{\varphi_{\ell+1}}(Z) T R_2^T\|_2 \quad (3.20)$$

as an approximation to the absolute 2-condition number  $\text{cond}_{\text{abs}}(\varphi_\ell, A)$ . And the relative 2-condition number  $\text{cond}_{\text{rel}}(\varphi_\ell, A)$  can be approximated by

$$\text{cond}_{\text{rel}}^{\text{II}}(\varphi_\ell, \tilde{A}) = \frac{\|A\|_2 \|R_1 L_{\varphi_{\ell+1}}(Z) T R_2^T\|_2}{\|R_1 [\varphi_{\ell+1}(Z)T] R_2^T\|_2}. \quad (3.21)$$

Similar to Strategy I, there is no need to form and store  $Q_1$  and  $Q_2$ , and the key is to evaluate 2-norms of some  $r$ -by- $r$  matrices.

**4. Numerical experiments.** In this section, we perform some numerical experiments to illustrate the numerical behavior of our new method. All the numerical experiments were run on a Dell PC with eight cores Intel(R) Core(TM)i7-2600 processor with CPU 3.40 GHz and RAM 16.0 GB, under the Windows 7 with 64-bit operating system. All the numerical results were obtained from MATLAB R2015b implementations with machine precision  $\epsilon \approx 2.22 \times 10^{-16}$ .

In all the examples, the sparse column-row approximation of  $A$  is computed by using the MATLAB functions `scra.m` and `spqr.m` due to G.W. Stewart<sup>1</sup>, where the tolerance `tol` is taken as  $\epsilon_{\text{col}} = \epsilon_{\text{row}} = 10^{-5}$ . In order to estimate the rank of a matrix, we consider the structural rank of  $A$ , i.e., `sprank(A)` that is obtained from running the MATLAB built-in function `sprank.m`. The matrix exponential is calculated by using the MATLAB built-in function `expm.m`, while the  $\varphi_\ell(\ell \geq 1)$  functions are computed by using the `phipade.m` function of the MATLAB package EXPINT [4].

<sup>1</sup><ftp://ftp.cs.umd.edu/pub/stewart/reports/Contents.html>.

**4.1. An application to data dimensionality reduction.** In this example, we show efficiency of our new method for computing matrix exponentials of large scale and low-rank matrices. Many data mining problems involve data sets represented in very high-dimensional spaces. In order to handle high dimensional data, the dimensionality needs to be reduced. Linear discriminant analysis (LDA) is one of notable subspace transformation methods for dimensionality reduction [12]. LDA encodes discriminant information by maximizing the between-class scatter, and meanwhile minimizing the within-class scatter in the projected subspace. Let  $X = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]$  be a set of training samples in an  $n$ -dimensional feature space, and assume that the original data is partitioned into  $K$  classes as  $X = [X_1, X_2, \dots, X_K]$ . We denote by  $m_j$  the number of samples in the  $j$ -th class, and thus  $\sum_{j=1}^K m_j = m$ . Let  $\mathbf{c}_j$  be the centroid of the  $j$ -th class, and  $\mathbf{c}$  be the global centroid of the training data set. If we denote  $\mathbf{e}_j = [1, 1, \dots, 1]^T \in \mathbb{R}^{m_j}$ , then the within-class scatter matrix is defined as

$$S_W = \sum_{j=1}^K \sum_{\mathbf{a}_i \in X_j} (\mathbf{a}_i - \mathbf{c}_j)(\mathbf{a}_i - \mathbf{c}_j)^T = H_W H_W^T,$$

where  $H_W = [X_1 - \mathbf{c}_1 \mathbf{e}_1^T, \dots, X_K - \mathbf{c}_K \mathbf{e}_K^T] \in \mathbb{R}^{n \times m}$ . The between-class scatter matrix is defined as

$$S_B = \sum_{j=1}^K n_j (\mathbf{c}_j - \mathbf{c})(\mathbf{c}_j - \mathbf{c})^T = H_B H_B^T,$$

where  $H_B = [\sqrt{n_1}(\mathbf{c}_1 - \mathbf{c}), \sqrt{n_2}(\mathbf{c}_2 - \mathbf{c}), \dots, \sqrt{n_K}(\mathbf{c}_K - \mathbf{c})] \in \mathbb{R}^{n \times K}$ . The LDA method is realized by maximizing the between-class scatter distance while minimizing the total scatter distance, and the optimal projection matrix can be obtained from solving the following large scale generalized eigenproblem

$$S_B \mathbf{x} = \lambda S_W \mathbf{x}. \quad (4.1)$$

However, the dimension of real data usually exceeds the number of training samples in practice (i.e.,  $n \gg m$ ), which results in  $S_W$  and  $S_B$  being singular. Indeed, suppose that the training vectors are linearly independent, then the rank of  $S_B$  and  $S_W$  is  $K - 1$  and  $m - K$ , respectively, which is much smaller than the dimensionality  $n$  [12]. This is called the small-sample-size (SSS) or undersampled problem [12, 31]. It is an intrinsic limitation of the classical LDA method, and is also a common problem in classification applications [31]. In other words, the SSS problem stems from generalized eigenproblems with singular matrices. So as to cure this drawback, a novel method based on matrix exponential, called exponential discriminant analysis method (EDA), was proposed in [44]. Instead of (4.1), the EDA method solve the following *generalized matrix exponential eigenproblem* [44]

$$\exp(S_B) \mathbf{x} = \lambda \exp(S_W) \mathbf{x}. \quad (4.2)$$

The EDA method is described as follows, for more details, refer to [44].

**ALGORITHM 2.** [44] **The exponential discriminant analysis method (EDA)**

**Input:** The data matrix  $X = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m] \in \mathbb{R}^{n \times m}$ , where  $\mathbf{a}_j$  represents the  $j$ -th training image.

**Output:** The projection matrix  $V$ .

1. Compute the matrices  $S_B$ ,  $S_W$ ,  $\exp(S_B)$ , and  $\exp(S_W)$ ;

2. Compute the eigenvectors  $\{\mathbf{x}_i\}$  and eigenvalues  $\{\lambda_i\}$  of  $\exp(S_W)^{-1}\exp(S_B)$ ;
3. Sort the eigenvectors  $V = \{\mathbf{x}_i\}$  according to  $\lambda_i$  in decreasing order;
4. Orthogonalize the columns of the projection matrix  $V$ .

As both  $\exp(S_W)$  and  $\exp(S_B)$  are symmetric positive definite (SPD), the difficulty of SSS problem can be cured naturally in the EDA method. The framework of the EDA method for dimensionality reduction has gained wide attention in recent years [1, 11, 40, 41, 43, 44]. However, the time complexity of EDA is dominated by the computation of  $\exp(S_B)$  and  $\exp(S_W)$ , which is prohibitively large as data dimension is large [44]. By Theorem 2.1, we can compute the large matrix exponentials as follows:

COROLLARY 4.1. *Under the above notations, we have that*

$$\exp(S_B) = I + H_B [\varphi_1(H_B^T H_B)] H_B^T, \quad (4.3)$$

and

$$\exp(S_W) = I + H_W [\varphi_1(H_W^T H_W)] H_W^T. \quad (4.4)$$

So we have the following algorithm for the matrix exponential discriminant analysis method.

ALGORITHM 3. **An algorithm for computing  $\exp(S_B)$  and  $\exp(S_W)$**

1. Given the data matrix  $X = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m] \in \mathbb{R}^{n \times m}$ , form  $H_B$  and  $H_W$ ;
2. Compute  $\varphi_1(H_B^T H_B)$  and  $\varphi_1(H_W^T H_W)$ ;
3. Store  $H_B, H_W$  and  $\varphi_1(H_B^T H_B), \varphi_1(H_W^T H_W)$  for  $\exp(S_B)$  and  $\exp(S_W)$ . If desired, form  $\exp(S_B)$  and  $\exp(S_W)$  in terms of (4.3) and (4.4).

Note that both  $H_B$  and  $H_W$  are already available in Step 1, so there is no need to perform rank-revealing decompositions to  $S_B$  and  $S_W$ . As a result, the computation of the two  $n \times n$  matrix exponentials  $\exp(S_B), \exp(S_W)$  reduces to that of  $\varphi_1(H_B^T H_B) \in \mathbb{R}^{K \times K}$  and  $\varphi_1(H_W^T H_W) \in \mathbb{R}^{m \times m}$ , with  $K, m \ll n$ .

Next we illustrate the efficiency of Algorithm 3 for the matrix exponential discriminant analysis method. There are three real-world databases in this example. The first one is the **ORL** database<sup>2</sup> that contains 400 face images of 40 individuals, and the original image size is  $92 \times 112 = 10304$ . The second test set is the **Yale** face database taken from the Yale Center for Computational Vision and Control<sup>3</sup>. It contains 165 grayscale images of  $K = 15$  individuals. The original image size is  $320 \times 243 = 77760$ . The third test set is the **Extended YaleB** database<sup>4</sup>. This database contains 5760 single light source images of 10 subjects, each seen under 576 viewing conditions. A subset of 38 classes with 2432 images are used in this example, 64 images of per individual with illumination.

In the **ORL** database, the images are aligned based on eye coordinates and are cropped and scaled to  $n = 32 \times 32$  and  $64 \times 64$ , respectively; and the original image size with  $n = 92 \times 112$  is also considered. In the **Yale** and the **Extended YaleB** databases, all images are aligned based on eye coordinates and are cropped and scaled to  $n = 32 \times 32$ ,  $64 \times 64$  and  $100 \times 100$ , respectively. In this example, a random subset with 3 images per subject is taken to form the training set, and the rest of the images are used as the testing set. Each column of the data matrices is scaled by its 2-norm.

<sup>2</sup><http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.

<sup>3</sup>[http://vision.ucsd.edu/datasets/yale\\_face\\_dataset\\_original/yalefaces.zip](http://vision.ucsd.edu/datasets/yale_face_dataset_original/yalefaces.zip).

<sup>4</sup><http://vision.ucsd.edu/~leekc/ExtYaleDatabase/Yale%20Face%20Database.html>.

In Algorithm 3, the CPU time consists of computing  $H_B, H_W$ , evaluating  $\varphi_1(H_B^T H_B)$  and  $\varphi_1(H_W^T H_W)$ , as well as forming  $\exp(S_B)$  and  $\exp(S_W)$  in terms of (4.3) and (4.4). In the original EDA algorithm (Algorithm 2), the CPU time consists of forming  $H_B, H_W$ , and the computation of  $\exp(S_B)$  and  $\exp(S_W)$  by using the MATLAB built-in function `expm.m`.

Let  $\widetilde{\exp(S_B)}, \widetilde{\exp(S_W)}$  be the “exact solutions” obtained from running `expm.m`, and let  $\exp(S_B), \exp(S_W)$  be the approximations obtained from (4.3) and (4.4). In this example, we define

$$\mathbf{Err}_B = \frac{\|\exp(S_B) - \widetilde{\exp(S_B)}\|_F}{\|\exp(S_B)\|_F}, \quad \mathbf{Err}_W = \frac{\|\exp(S_W) - \widetilde{\exp(S_W)}\|_F}{\|\exp(S_W)\|_F},$$

as the relative errors of the approximations  $\widetilde{\exp(S_B)}, \widetilde{\exp(S_W)}$ , respectively, and denote by

$$\mathbf{Rel\_ErrF} = \max(\mathbf{Err}_B, \mathbf{Err}_W)$$

the maximal value of the two relative errors. Table 1 lists the CPU time in seconds of Algorithm 3, `expm.m`, and the values of the maximal relative errors  $\mathbf{Rel\_ErrF}$ .

Database	$n$	Algorithm 3	<code>expm.m</code>	$\mathbf{Rel\_ErrF}$
ORL	1024	0.08	0.26	$2.20 \times 10^{-15}$
	4096	0.26	17.3	$3.22 \times 10^{-15}$
	10304	1.28	261.1	$4.49 \times 10^{-15}$
Yale	1024	0.08	0.25	$2.11 \times 10^{-15}$
	4096	0.24	17.3	$3.10 \times 10^{-15}$
	10000	1.13	238.5	$4.35 \times 10^{-15}$
Extended YaleB	1024	0.08	0.27	$2.13 \times 10^{-15}$
	4096	0.27	17.3	$3.14 \times 10^{-15}$
	10000	1.22	238.6	$4.50 \times 10^{-15}$

Example 1, Table 1: CPU time in seconds and the relative errors for computing  $\exp(S_B)$  and  $\exp(S_W)$ .

We observe from Table 1 that Algorithm 3 runs much faster than `expm.m`, especially when  $n$  is large. For instance, when the dimensionality of the datasets is around  $10^4$ , `expm.m` requires about 240 seconds, while our new method only needs about 1.2 seconds, a great improvement. Furthermore, the relative errors of our approximations are in the order of  $\mathcal{O}(10^{-15})$ , implying that our new method is numerically stable. Thus, the new method is very efficient and reliable for solving large matrix exponential problems arising in the EDA framework for high dimensionality reduction.

**4.2. Computing  $\varphi$ -functions of matrices with low rank or fast decaying singular values.** In this example, we show the efficiency of Algorithm 1 for *consecutively* computing several  $\varphi$ -functions of  $A$  with low rank or with fast decaying singular values. The test matrices are available from [10, 39], and Table 2 lists problem characteristics of these matrices. Here the first five matrices are rank-deficient while the last three are full rank but with fast decaying singular values.

In this example, we compare Algorithm 1 with `expm.m/philpade.m`, that is, `expm.m` for the matrix exponential  $\exp(A)$  and `philpade.m` for  $\varphi_\ell(A)$ ,  $\ell = 1, 2, 3, 4$ . In Algorithm 1, the CPU time consists of computing the sparse column-row approximation

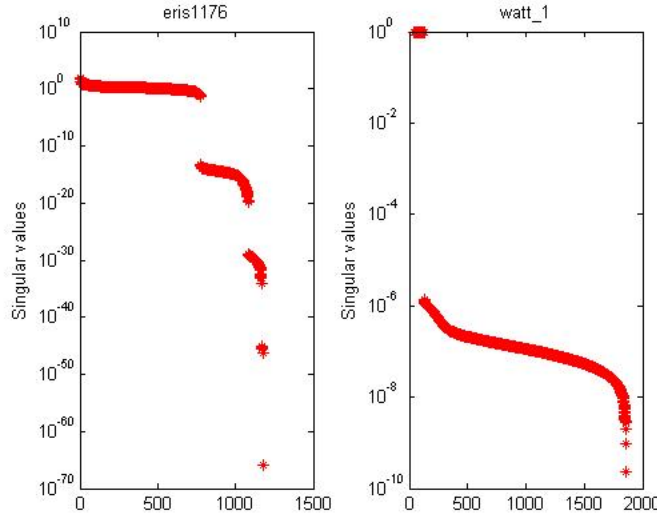
(SCR), the evaluation of  $\varphi_\ell(Z)$  ( $\ell = 1, 2, 3, 4, 5$ ) by using `phipade.m`, as well as forming  $\varphi_\ell(A)$  in terms of (2.2),  $\ell = 0, 1, 2, 3, 4$ . In `expm.m/phipade.m`, the CPU time consists of computing  $\varphi_\ell(A)$  by using `expm.m` ( $\ell = 0$ ) and `phipade.m`,  $\ell = 1, 2, 3, 4$ . In order to measure the accuracy of the computed solutions, we define the maximal relative error as

$$\mathbf{Rel\_ErrF} = \max_{0 \leq \ell \leq 4} \frac{\|\varphi_\ell(A) - \widetilde{\varphi_\ell(A)}\|_F}{\|\varphi_\ell(A)\|_F},$$

where  $\varphi_\ell(A)$  is the “exact solution” obtained from `expm.m` as  $\ell = 0$  and `phipade.m` as  $\ell = 1, 2, 3, 4$ ; and  $\widetilde{\varphi_\ell(A)}$  is the approximation obtained from running Algorithm 1. Table 3 lists the numerical results.

Test matrix	$n$	$\text{sprank}(A)$	$\text{nnz}(A)$	Description
man5976	5976	5882	225046	Structural problem
Movies	5757	1275	24451	Directed network
lock3491	3491	3416	160444	Structural problem
cegb3306	3306	3222	74916	Finite element framework
zenios	2873	266	1314	Optimization problem
watt_1	1856	1856	11360	Computational fluid dynamics
watt_2	1856	1856	11550	Computational fluid dynamics
eris1176	1176	1176	18552	Power network problem

Example 2, Table 2: Problem characteristics of the test matrices, where “ $\text{nnz}(A)$ ” denotes the number of nonzero elements of  $A$ .



Example 2, Figure 1: Singular values of the `eris1176` matrix and the `watt_1` matrix.

It is seen from Table 3 that Algorithm 1 works quite well, and we provide a competitive candidate for consecutively computing several  $\varphi$ -functions of large sparse matrices with low rank or with fast decaying singular values. Firstly, Algorithm 1 runs much faster than `expm.m/phipade.m`. For instance, 205.5 seconds vs. 10314 seconds

Test matrix	Algorithm 1	expm.m/hipade.m	Rel.ErrF
man5976*	205.5	10314.0	$1.10 \times 10^{-12}$
Movies	360.6	456.1	$4.32 \times 10^{-14}$
lock3491*	5.49	2469.6	$5.12 \times 10^{-13}$
cegb3306*	3.11	1599.7	$1.24 \times 10^{-13}$
zenios	0.63	3.63	$1.17 \times 10^{-7}$
watt_1	0.19	30.4	$1.93 \times 10^{-7}$
watt_2	0.19	39.1	$2.02 \times 10^{-7}$
eris1176*	6.74	85.7	$2.86 \times 10^{-12}$

Example 2, Table 3: CPU time in seconds and the maximal relative error for computing  $\varphi_\ell$  matrix functions,  $\ell = 0, 1, 2, 3, 4$ ; where “\*” denotes we compute  $\varphi_\ell(-A)$  instead of  $\varphi_\ell(A)$ .

for the **man5976** matrix, 5.49 seconds vs. 2469.6 seconds for the **lock3491** matrix, and 3.11 seconds vs. 1599.7 seconds for the **cegb3306** matrix. Secondly, the accuracy of our approximations is satisfactory in most cases. However, for the **zenios**, **watt\_1** and **watt\_2** matrices, the relative errors **Rel.ErrF** are in the order of  $\mathcal{O}(10^{-7})$ . In Figure 1, we plot the singular values of the **eris1176** matrix and the **watt\_1** matrix. It is observed that the **eris1176** matrix has faster decaying singular values, while the decaying speed of the singular values of the **watt\_1** matrix is relatively slower. Indeed, the error  $\|A - XTY^T\|_F$  from the SCR decomposition, with respect to the three matrices **zenios**, **watt\_1** and **watt\_2** are about  $7.75 \times 10^{-6}$ ,  $9.97 \times 10^{-6}$  and  $9.98 \times 10^{-6}$ , respectively, while that of the **eris1176** matrix is about  $2.49 \times 10^{-10}$ . In terms of Theorem 2.4, the accuracy of the computed solution of the **eris1176** matrix can be higher than that of **zenios**, **watt\_1** and **watt\_2**, provided that the condition numbers are comparable. Thus, our new method is suitable to  $\varphi$ -functions of large matrices with low rank or with fast decaying singular values.

**4.3. Estimating the relative and absolute condition numbers.** In this example, we demonstrate the efficiency of Strategy I and Strategy II for estimating the absolute and relative condition numbers of  $\varphi$ -functions. There are four test matrices in this example, which are available from [10, 39]. The problem characteristics of these matrices are given in Table 4.

We compare Strategy I and Strategy II with the **funm\_condest1.m** function in the Matrix Function Toolbox [46]. In Strategy I and Strategy II, the matrices  $R_1, R_2$  are obtained from the QR decompositions of  $X$  and  $Y$ , respectively, by the MATLAB built-in function **qr.m**. The matrix  $L_{\varphi_{\ell+1}}(Z)$  in (3.21) is calculated by using the **funm\_condest\_fro.m** function in the Matrix Function Toolbox. When  $\ell = 0$ , the parameter “fun” in **funm\_condest1.m** is called by the MATLAB built-in function **expm.m**, while  $\ell > 0$  this parameter is called by the **hipade.m** function in the EXPINT package. The CPU time for both Strategy I and Strategy II is composed of computing the sparse column-row approximation, and calculating (3.10), (3.13) or (3.20), (3.21), respectively. In (3.13) and (3.21),  $\|A\|_2$  is evaluated by using the MATLAB built-in function **svds.m**. Tables 5–8 report the numerical results, where “**Relative.Est**” and “**Absolute.Est**” denote an estimation to the relative and the absolute condition number, respectively.

As was pointed out in [20, pp.64], for the absolute and relative condition numbers, “what is needed is an estimate that is of the correct order of magnitude in practice—more than one correct significant digit is not needed”. Recall that the **funm\_condest1.m** function estimates the 1-norm relative and absolute condition num-



Network matrix	$n$	$\text{sprank}(A)$	$\text{nnz}(A)$
California	9664	1686	16150
EVA	8497	1303	6726
EPA	4772	986	8965
eris1176	1176	1176	18552

Example 3, Table 4: Problem characteristics of the test matrices, where “ $\text{nnz}(A)$ ” denotes the number of nonzero elements of  $A$ .

bers, while Strategy I and Strategy II estimate the 2-norm condition numbers. Compared with the numerical results of `funm_condest1`, we see from Tables 5–8 that both Strategy I and Strategy II capture the correct order of magnitude of the condition numbers in many cases, and we can not tell which one, Strategy I or Strategy II, is *definitely* better than the other. We find that Strategy I runs (a little) faster than Strategy II in terms of CPU time. The reason is that we have to evaluate  $L_{\varphi_{\ell+1}}(Z)$  iteratively via the `funm_condest_fro.m` function.

On the other hand, it is observed from the numerical results that our new strategies often run much faster than `funm_condest1`. For instance, as  $\ell = 0$ , the new methods used 752.0 and 764.2 seconds for the **California** matrix, respectively, while `funm_condest1` used 1477.5 seconds. As  $\ell = 1$ , the new methods used 757.9 and 788.7 seconds, respectively, while `funm_condest1` used 5266.8 seconds. The improvement is impressive. Specifically, as  $\ell \geq 2$ , for some large matrices such as **California** and **EVA**, `funm_condest1` fails to converge within 3 hours. As a comparison, our new methods work quite well. Thus, we benefit from our new strategies, and provide competitive alternatives for estimating the relative and absolute condition numbers of  $\varphi$ -functions with respect to large sparse matrices.

$\ell$	Method	Relative_Est	Absolute_Est	CPU
0	<code>funm_condest1</code>	$6.82 \times 10^3$	$7.53 \times 10^4$	1477.5
	Strategy I	$4.01 \times 10^2$	$3.92 \times 10^4$	752.0
	Strategy II	21.8	$2.13 \times 10^3$	764.2
1	<code>funm_condest1</code>	$1.33 \times 10^4$	$1.98 \times 10^4$	5266.8
	Strategy I	$6.53 \times 10^2$	$8.58 \times 10^3$	757.9
	Strategy II	25.1	$3.29 \times 10^2$	788.7
2	<code>funm_condest1</code>	—	—	>3h
	Strategy I	$1.38 \times 10^3$	$2.41 \times 10^3$	758.1
	Strategy II	15.6	27.2	801.0
3	<code>funm_condest1</code>	—	—	>3h
	Strategy I	$2.57 \times 10^3$	$5.80 \times 10^2$	757.9
	Strategy II	34.7	7.82	827.0
4	<code>funm_condest1</code>	—	—	>3h
	Strategy I	$4.10 \times 10^3$	$1.15 \times 10^2$	761.1
	Strategy II	82.6	2.31	852.1

Example 3, Table 5: Estimation of the relative and absolute condition numbers of  $\varphi_\ell(A)$ ,  $\ell = 0, 1, 2, 3, 4$ , and the CPU time in seconds, where “>3h” denotes the algorithm fails to converge within 3 hours. The **California** matrix,  $n = 9664$ ,  $\text{sprank}(A) = 1686$ .

$\ell$	Method	Relative_Est	Absolute_Est	CPU
0	<b>funm_condest1</b>	$7.74 \times 10^3$	$1.83 \times 10^4$	700.0
	Strategy I	$3.53 \times 10^2$	$4.37 \times 10^2$	412.0
	Strategy II	$1.07 \times 10^3$	$1.32 \times 10^3$	420.1
1	<b>funm_condest1</b>	$7.07 \times 10^3$	$7.22 \times 10^3$	5846.1
	Strategy I	$3.17 \times 10^2$	$1.59 \times 10^2$	419.4
	Strategy II	$5.38 \times 10^2$	$2.69 \times 10^2$	425.2
2	<b>funm_condest1</b>	—	—	>3h
	Strategy I	$2.72 \times 10^2$	45.3	415.2
	Strategy II	$3.21 \times 10^2$	53.5	429.8
3	<b>funm_condest1</b>	—	—	>3h
	Strategy I	$2.49 \times 10^2$	10.4	408.9
	Strategy II	$1.22 \times 10^2$	5.08	440.9
4	<b>funm_condest1</b>	—	—	>3h
	Strategy I	$2.36 \times 10^2$	1.97	408.0
	Strategy II	73.3	0.61	447.1

Example 3, Table 6: Estimation of the relative and absolute condition numbers of  $\varphi_\ell(A)$ ,  $\ell = 0, 1, 2, 3, 4$ , and the CPU time in seconds, where “>3h” denotes the algorithm fails to converge within 3 hours. The EVA matrix,  $n = 8497$ ,  $\text{sprank}(A)=1303$ .

$\ell$	Method	Rel_Est	Abs_Est	CPU
0	<b>funm_condest1</b>	$9.90 \times 10^3$	$2.60 \times 10^4$	90.1
	Strategy I	$2.48 \times 10^2$	$1.12 \times 10^3$	90.6
	Strategy II	$3.34 \times 10^4$	$1.52 \times 10^5$	95.4
1	<b>funm_condest1</b>	$1.09 \times 10^4$	$8.05 \times 10^3$	272.1
	Strategy I	$2.51 \times 10^2$	$3.06 \times 10^2$	90.3
	Strategy II	$2.25 \times 10^4$	$2.74 \times 10^4$	100.6
2	<b>funm_condest1</b>	$8.95 \times 10^3$	$2.03 \times 10^3$	443.7
	Strategy I	$2.65 \times 10^2$	76.9	90.8
	Strategy II	$1.47 \times 10^4$	$4.27 \times 10^3$	105.5
3	<b>funm_condest1</b>	$7.86 \times 10^3$	$4.25 \times 10^2$	773.0
	Strategy I	$2.85 \times 10^2$	17.2	92.5
	Strategy II	$9.71 \times 10^3$	$5.87 \times 10^2$	110.7
4	<b>funm_condest1</b>	$7.17 \times 10^3$	75.2	1357.7
	Strategy I	$3.03 \times 10^2$	3.29	91.7
	Strategy II	$5.29 \times 10^3$	57.6	116.7

Example 3, Table 7: Estimation of the relative and absolute condition numbers of  $\varphi_\ell(A)$ ,  $\ell = 0, 1, 2, 3, 4$ , and the CPU time in seconds. The EPA matrix,  $n = 4772$ ,  $\text{sprank}(A)=986$ .

**4.4. Sharpness of Theorem 2.4.** In this example, we aim to show the sharpness of Theorem 2.4. The test matrix is the **watt\_1** matrix used in Example 2; see Table 2. It is a  $1856 \times 1856$  full-rank matrix with fast decaying singular values. So as to show the sharpness of Theorem 2.4, we denote by

$$\mathbf{Abs\_Err2} = \|\varphi_\ell(A) - \varphi_\ell(\tilde{A})\|_2,$$

$\ell$	Method	Relative_Est	Absolute_Est	CPU
0	<code>funm_condest1</code>	575.1	$1.45 \times 10^3$	2.47
	Strategy I	$1.10 \times 10^4$	$1.88 \times 10^4$	7.33
	Strategy II	$2.72 \times 10^3$	$4.66 \times 10^3$	11.1
1	<code>funm_condest1</code>	$1.09 \times 10^3$	548.7	38.2
	Strategy I	$1.10 \times 10^4$	$3.69 \times 10^3$	7.48
	Strategy II	$2.66 \times 10^3$	890.5	14.5
2	<code>funm_condest1</code>	$1.75 \times 10^3$	171.0	60.2
	Strategy I	$1.10 \times 10^4$	679.0	7.78
	Strategy II	$2.71 \times 10^3$	167.6	17.6
3	<code>funm_condest1</code>	$2.37 \times 10^3$	42.2	82.2
	Strategy I	$1.10 \times 10^4$	114.9	8.09
	Strategy II	$2.82 \times 10^3$	29.4	20.5
4	<code>funm_condest1</code>	$2.85 \times 10^3$	8.49	103.4
	Strategy I	$1.10 \times 10^4$	17.6	8.26
	Strategy II	$2.88 \times 10^3$	4.61	24.2

Example 3, Table 8: Estimation of the relative and absolute condition numbers of  $\varphi_\ell(-A)$ ,  $\ell = 0, 1, 2, 3, 4$ , and the CPU time in seconds. The `eris1176` matrix,  $n = 1176$ , `sprank(A)=1176`.

and

$$\mathbf{Rel\_Err2} = \frac{\|\varphi_\ell(A) - \varphi_\ell(\tilde{A})\|_2}{\|\varphi_\ell(A)\|_2},$$

the absolute and relative errors of the computed solutions  $\varphi_\ell(\tilde{A})$  with respect to  $\varphi_\ell(A)$  in terms of 2-norm. The values of  $\text{cond}_{\text{abs}}(\varphi_\ell, A)$  and  $\text{cond}_{\text{rel}}(\varphi_\ell, A)$  in the upper bounds of (2.15) and (2.16) are estimated by Strategy I or Strategy II, respectively, and the corresponding estimations are denoted by “(2.15)–StrI”, “(2.16)–StrI”, and “(2.15)–StrII”, “(2.16)–StrII”, respectively. Table 9 lists the numerical results.

We see from Table 9 that both (2.15) and (2.16) are very sharp, which justify Strategy I and Strategy II for estimating  $\text{cond}_{\text{abs}}(\varphi_\ell, A)$  and  $\text{cond}_{\text{rel}}(\varphi_\ell, A)$ . We find that the values of (2.15)–StrII and (2.16)–StrII are a little smaller than those of `Abs_Err2` and `Rel_Err2` in many cases. In fact, both Strategy I and Strategy II only give approximations to the absolute and relative condition numbers, which are neither upper bounds nor lower bounds theoretically.

**5. Concluding remarks.** In this paper we consider the computations, error analysis, implementations and applications of  $\varphi$ -functions for large sparse matrices with low rank or with fast decaying singular values. Given a sparse column-row approximation of  $A$ , we take into account how to compute the matrix function series  $\varphi_\ell(A)$  ( $\ell = 0, 1, 2, \dots, p$ ) efficiently, and to estimate their 2-norm Fréchet relative and absolute condition numbers effectively.

The numerical behavior of our new method is closely related to that of reduced-rank approximation of large sparse matrices [8, 36]. Thus, a promising research area is to seek new technologies to improve the performance of the sparse column-row approximation algorithm on very large matrices. Another interesting topic is to combine other advanced algorithms such as the randomized singular value decomposition algorithm [17, 28] with our new strategies for the computation of functions of large sparse matrices.

	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
<b>Abs_Err2</b>	$1.07 \times 10^{-6}$	$5.35 \times 10^{-7}$	$1.78 \times 10^{-7}$	$4.45 \times 10^{-8}$	$8.91 \times 10^{-9}$
<b>Rel_Err2</b>	$3.93 \times 10^{-7}$	$3.11 \times 10^{-7}$	$2.48 \times 10^{-7}$	$2.04 \times 10^{-7}$	$1.73 \times 10^{-7}$
(2.15)–StrI	$3.16 \times 10^{-6}$	$1.32 \times 10^{-6}$	$4.01 \times 10^{-7}$	$9.48 \times 10^{-8}$	$1.83 \times 10^{-8}$
(2.16)–StrI	$1.84 \times 10^{-6}$	$1.84 \times 10^{-6}$	$1.84 \times 10^{-6}$	$1.84 \times 10^{-6}$	$1.84 \times 10^{-6}$
(2.15)–StrII	$1.83 \times 10^{-7}$	$5.24 \times 10^{-8}$	$1.19 \times 10^{-8}$	$2.22 \times 10^{-9}$	$3.47 \times 10^{-10}$
(2.16)–StrII	$1.06 \times 10^{-7}$	$7.30 \times 10^{-8}$	$5.43 \times 10^{-8}$	$4.30 \times 10^{-8}$	$3.49 \times 10^{-8}$

Example 4, Table 9: Absolute and relative errors and their estimations,  $\ell = 0, 1, 2, 3, 4$ . The `watt_1` matrix, with  $\|A - \tilde{A}\|_2 \approx 1.07 \times 10^{-6}$ . Here (2.15)–StrI, (2.16)–StrI, (2.15)–StrII, (2.16)–StrII denote the values of  $\text{cond}_{\text{abs}}(\varphi_\ell, A)$  and  $\text{cond}_{\text{rel}}(\varphi_\ell, A)$  in the upper bounds of (2.15) and (2.16), are estimated by using Strategy I and Strategy II, respectively.

**Acknowledgments.** We would like to thank Juan-juan Tian for helpful discussions.

#### REFERENCES

- [1] N. AHMED, *Exponential discriminant regularization using nonnegative constraint and image descriptor*, IEEE 9th International Conference on Emerging Technologies, pp.1–6, 2013.
- [2] A. AL-MOHY AND N.J. HIGHAM, *Computation the action of the matrix exponential, with an application to exponential integrators*, SIAM J. Sci. Comput., 33 (2011), pp. 488–511.
- [3] M. BENZI, E. ESTRADA, AND C. KLYMKO, *Ranking hubs and authorities using matrix functions*, Linear Algebra Appl., 438 (2013), pp. 2447–2474.
- [4] H. BERLAND, B. SKAFLESTAD, AND W. WRIGHT, *EXPINT–A matlab package for exponential integrators*, ACM Tran. Math. Soft., 33 (2007), Article 4.
- [5] M. BERRY AND M. BROWNE, *Understanding Search Engines: Mathematical Modeling and Text Retrieval*, SIAM, Philadelphia, 1999.
- [6] M. BERRY, Z. DRMAČ, AND E. JESSUP, *Matrices, vector spaces, and information retrieval*, SIAM Rev., 41 (1999), pp. 335–362.
- [7] M. BERRY, S. DUMAIS, AND G. O’BRIEN, *Using linear algebra for intelligent information retrieval*, SIAM Rev., 37 (1995), pp. 573–595.
- [8] M. BERRY, S. PULATOVA, AND G.W. STEWART, *Computing sparse reduced-rank approximations to sparse matrices*, ACM Tran. Math. Soft., 31 (2005), pp. 252–269.
- [9] G. BEYLKIN, J. KEISER, AND L. VOZOVOL, *A new class of time discretization schemes for the solution of nonlinear PDEs*, J. Comput. Phys., 147 (1998), pp. 362–387.
- [10] T. DAVIS AND Y. HU, *The University of Florida sparse matrix collection*, ACM Tran. Math. Soft., 38 (2011), Article 1. <http://www.cise.ufl.edu/research/sparse/matrices/>.
- [11] F. DORNAIKA, A. BOSAGHZADEH, *Exponential local discriminant embedding and its application to face recognition*, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 43 (2013), pp. 921–934.
- [12] R. DUDA, P. HART, AND D. STORK, *Pattern Classification*, 2nd ed. New York: Wiley, 2000.
- [13] E. ESTRADA AND N. HATANO, *Communicability in complex networks*, Physical Review E, 77: 036111, 2008.
- [14] E. ESTRADA AND D.J. HIGHAM, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 671–696.
- [15] E. ESTRADA AND J. RODRÍGUEZ-VELÁZQUEZ, *Subgraph centrality in complex networks*, Physical Review E, 71: 056103, 2005.
- [16] G.H. GOLUB AND C.F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, USA, Forth edition, 2013.
- [17] N. HALKO, P. MARTINSSON, AND J. TROPP, *Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288.
- [18] P. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
- [19] N.J. HIGHAM, *Accuracy and stability of numerical algorithms*, second edition, SIAM, Philadelphia, 2002.

- [20] N.J. HIGHAM, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.
- [21] N.J. HIGHAM, *The scaling and squaring method for the matrix exponential revisited*, SIAM Matrix Anal. Appl., 51 (2009), pp. 747–764.
- [22] M. HOCHBRUCK, C. LUBICH, AND H. SELHOFER, *Exponential integrators for large systems of differential equations*, SIAM J. Sci. Comput., 19 (1998), pp. 1552–1574.
- [23] M. HOCHBRUCK AND A. OSTERMANN, *Exponential integrators*, Acta Numer., pp. 209–286, 2010.
- [24] B. HOFMANN, *Regularization for Applied Inverse and Ill-posed Problems*, Teubner, Stuttgart, German, 1986.
- [25] P. JIANG AND M. BERRY, *Solving total least squares problems in information retrieval*, Linear Algebra Appl., 316 (2000), pp. 137–156.
- [26] A. KASSAM AND L.N. TREFETHEN, *Fourth-order time-stepping for stiff PDEs*, SIAM J. Sci. Comput., 26 (2005), pp. 1214–1233.
- [27] Y. LU, *Computing a matrix function for exponential integrators*, J. Comput. Appl. Math., 161 (2003), pp. 203–216.
- [28] M. MAHONEY, *Randomized algorithms for matrices and data*, Foundations and Trends in Machine Learning, NOW Publishers, Volume 3, Issue 2, 2011.
- [29] C. MOLER AND C.F. VAN LOAN, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*, SIAM Rev., 45 (2003), pp. 3–49.
- [30] J. NIESEN AND W. WRIGHT, *Algorithm 919: A Krylov subspace algorithm for evaluating the  $\varphi$ -functions appearing in exponential integrators*, ACM Trans. Math. Soft., 38 (2012), Article 22.
- [31] C. PARK AND H. PARK, *A comparison of generalized linear discriminant analysis algorithms*, Pattern Recognition, 41 (2008), pp. 1083–1097.
- [32] J. RICE, *A theory of condition*, SIAM J. Numer. Anal., 3 (1966), pp. 287–310.
- [33] T. SCHMELZER AND L.N. TREFETHEN, *Evaluating matrix functions for exponential integrators via Carathéodory-Fejér approximation and contour integrals*, Electron. Trans. Numer. Anal., 29 (2007), pp. 1–18.
- [34] R. SIDJE, *EXPOKIT: Software package for computing matrix exponentials*, ACM Tran. Math. Soft., 24 (1998), pp. 130–156.
- [35] B. SKAFLESTAD AND W. WRIGHT, *The scaling and squaring method for matrix functions related to the exponential*, Appl. Numer. Math., 59 (2009), pp. 783–799.
- [36] G.W. STEWART, *Four algorithms for the efficient computation of truncated pivoted qr approximations to a sparse matrix*, Numer. Math., 83 (1999), pp. 313–323.
- [37] G.W. STEWART, *Error analysis of the quasi-Gram-Schmidt algorithm*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 493–506.
- [38] G. STUART, M. BERRY, *A comprehensive whole genome bacterial phylogeny using correlated peptide motifs defined in a high dimensional vector space*, J. Bioinformatics and Computational Bio., 1 (2003), pp. 475–493.
- [39] P. TSPPRAS, *Datasets for Experiments on Link Analysis Ranking Algorithms*, <http://www.cs.toronto.edu/~tsap/experiments/datasets/index.html>.
- [40] S. WANG, H. CHEN, X. PENG, AND C. ZHOU, *Exponential locality preserving projections for small sample size problem*, Neurocomputing, 74 (2011), pp. 36–54.
- [41] S. WANG, S. YAN, J. YANG, C. ZHOU, AND X. FU, *A general exponential framework for dimensionality reduction*, IEEE Tran. Image Process., 23 (2014), pp. 920–930.
- [42] G. WU, L. ZHANG, AND T. XU, *A framework of the harmonic Arnoldi method for evaluating  $\varphi$ -functions with applications to exponential integrators*, Adv. Comput. Math., 42(2016), pp. 505–541.
- [43] L. YAN AND J. PAN, *Two-dimensional exponential discriminant analysis and its application to face recognition*, International Conference on Computational Aspects of Social Networks (CAsoN), pp. 528–531, 2010.
- [44] T. ZHANG, B. FANG, Y. TANG, Z. SHANG, AND B. XU, *Generalized discriminant analysis: a matrix exponential approach*, IEEE Transactions on Systems Man and Cybernetics-part B: cybernetics, 40 (2010), pp. 186–197.
- [45] Z. ZHANG, H. ZHA, AND H. SIMON, *Low-rank approximations with sparse factors I: Basic algorithms and error analysis*, SIAM J. Matrix Anal. Appl., 23 (2002), pp. 706–727.
- [46] THE MATRIX FUNCTION TOOLBOX. <http://www.mathworks.com/matlabcentral/fileexchange/20820-the-matrix-function-toolbox>.